DOCUMENT RESUME

ED 407 415 TM 026 436

AUTHOR Minke, Amy
TITLE Conducting Repeated Measures Analyses: Experimental Design Considerations.
PUB DATE Jan 97
NOTE 17p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 23-25, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Experiments; *Multivariate Analysis; *Research Design; Research Methodology; Sample Size; *Statistical Significance; *Validity
IDENTIFIERS *Repeated Measures Design; Sphericity Tests; Univariate Analysis

ABSTRACT

Repeated measures experimental designs, often referred to as "within-subjects" designs, offer researchers opportunities to study research effects while "controlling" for subjects. These designs offer greater statistical power relative to sample size. However, threats to internal validity such as carryover or practice effects need to be taken into consideration. Once data are gathered, researchers have several options for data analysis. If univariate statistical methods are used, omnibus tests can be used, but they must be evaluated for violation of the sphericity assumption, or planned comparisons can be used. Researchers may also use multivariate statistical methods or they may implement both univariate and multivariate approaches while controlling for experiment-wise error. This paper considers both univariate and multivariate approaches to analyzing repeated measures design. Within the univariate discussion, analysis of variance and regression approaches are compared. Also, the assumptions necessary to perform statistical significance tests and how to investigate possible violations of the sphericity assumption are discussed. (Contains six tables and eight references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Running Head: Repeated Measures

ED 407 415

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Amy Minke
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Conducting Repeated Measures Analyses: Experimental Design Considerations

Amy Minke

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, Texas, January 23-25, 1997.

TM026436

BEST COPY AVAILABLE

Abstract

Repeated measures experimental designs, often referred to as within-subjects designs, offer researchers opportunities to study research effects while "controlling" for subjects. These designs offer greater statistical power relative to sample size. This paper considers both univariate and multivariate approaches to analyzing repeated measures data. Within the univariate discussion, ANOVA and regression approaches are compared. Also, the assumptions necessary to perform statistical significance tests and how to investigate possible violations of the sphericity assumption are discussed.

Experimental designs called "repeated measures" designs are characterized by having more than one measurement of at least one given variable for each subject. A well-known repeated measures design is the pretest, posttest experimental design, with intervening treatment; this design measures the same subjects twice on an intervally-scaled variable, and then uses the correlated or dependent samples t test in the analysis (Stevens, 1996). As another example, in a 2 × 3 repeated measures factorial design, each subject has a score for each of the combinations of the factors, or in each of the six cells of the data matrix (Huck & Cormier, 1996).

There are many research hypotheses that can be tested using repeated measures designs, such as hypotheses that compare the same subjects under several different treatments, or those that follow performance over time. Repeated measures designs are quite versatile, and researchers use many different designs and call the designs by many different names. For example, a one-way repeated measures ANOVA may be known as a one-factor within-subjects ANOVA, a treatments-by-subjects ANOVA, or a randomized blocks ANOVA. A two-way repeated measures ANOVA may be referred to as a two-way within-subjects ANOVA, a two-way ANOVA with repeated measures on both factors, a multiple treatments-by-subjects ANOVA, or treatments-by-treatments-by-subjects ANOVA (Huck & Cormier, 1996). There are also "mixed model" designs which use both "between" variables and "within" variables (Hertzog & Rovine, 1985).

In repeated measures designs, these terms differentiate among repeated and non-repeated factors. A "between" variable is a non-repeated or grouping factor, such as gender or experimental group, for which subjects will appear in only one

level. A "within" variable is a repeated factor for which subjects will participate in each level, e.g. subjects participate in both experimental conditions, albeit at different times (Stevens, 1996).

The primary benefit of a repeated measures design is statistical power relative to sample size which is important in many real-world research situations. Repeated measures designs use the same subjects throughout different treatments and thus, require fewer subjects overall. Because the subjects are constant, the variance due to subjects can be partitioned out of the error variance term, thereby making any statistical tests more powerful (Stevens, 1996).

Though the benefits of repeated measures designs can be great, there are internal validity issues that must be addressed. "Carryover" effects are effects from one treatment that may extend into and affect the next treatment. They may be effects such as tracking memory over time or investigating practice or fatigue on a targeted behavior. However, carryover effects may be detrimental to a study, for example if a second drug treatment is administered without the previous drug passing out of the subject's system (Edwards, 1985). This internal validity threat can be controlled through counterbalancing. By varying the presentation order of treatments, either randomly or systematically, interaction between treatment order and main effect can be investigated through data analysis (Huck & Cormier, 1996). However, even with couterbalancing, carryover effects can raise issues involving external validity.

There are several ways to approach repeated measures analyses. Edwards (1985) presented two heuristic examples of repeated measures analysis performed

through ANOVA and through regression. The following discussion will consider a one-way repeated measures design, but the concepts generalize to other designs. Table 1 represents a general data matrix for a one-way repeated measures design with n subjects and k treatments or repeated measures. Table 2 presents sample data from Edwards (1985). Tables 3 and 4 represent ANOVA summary tables for the general and example data matrices, respectively.

Insert Tables 1, 2, 3, & 4 about here

Notice how the general ANOVA table differs from a one-way independent samples ANOVA table; the row for Subjects acts as another factor and the residual or error term is the interaction between Subjects and Treatments. This difference arises because Subjects are constant throughout the treatments and thus subject effects may be partitioned out of the error variance. There is still only one effect of interest, Treatments, with only one test statistic (Huck & Cormier, 1996).

The same analysis may be performed through a regression rubric. First, define k-1 mutually orthogonal contrasts or vectors to represent the treatments. For the example, there are k=2 treatments, so there needs to be 2-1=1 "mutually orthogonal" vector to define the set. Treatment 1 is coded as 1 and Treatment 2 is coded as -1. Table 5 reports the resulting vector. Second, define n-1 mutually orthogonal vectors to represent the subjects. These n-1 subject vectors may be condensed into one vector consisting of the sum of the k scores from the repeated measures for each subject. Table 5 reports this vector, as well.

Insert Table 5 about here

The resulting set of k vectors, say (X_1, X_2 ... X_k) will be mutually orthogonal which implies, by definition, that $\Sigma X_i = 0$, $X_i \times X_j = 0$ and $r_{ij} = 0$. To find the squared correlation between any X_i and Y, the vector consisting of the scores on the repeated measures, we use the following formula:

$$r_{yi}^2 = \frac{(\Sigma xy)^2}{\Sigma x_i^2 \Sigma y^2}$$

Since $\Sigma X_i = 0$, $\Sigma x_i y = \Sigma X_i Y$ and $\Sigma x_i^2 = \Sigma X_i^2$. Then, the formula reduces to

$$r_{yi}^2 = \frac{(\Sigma X_i Y)^2}{\Sigma X_i^2 \Sigma Y^2}$$

Because the intercorrelations between the X_i are zero, the formula for the multiple

R^2 simplifies to $R^2 = \Sigma r_{yi}^2$

We know $R^2 = SS_{Ex}/TSS = (SS_T + SS_S)/TSS$

Thus, the residual is $(1 - R^2) = SS_{ST}/TSS$

The multiple correlation due to treatments is

$$R_T^2 = \Sigma_{(i=1 \text{ to } k-1)} r_{yi}^2$$

and $R_T^2 = SS_T/TSS$

from the ANOVA summary table. Thus, we have computed the equivalent effect size as found through ANOVA. We can now compute the omnibus F statistic:

$$F = \frac{R_T^2/(k-1)}{(1 - R^2)/(k-1)(n-1)}$$

with degrees of freedom k-1 and (k-1)(n-1). This test statistic is equivalent to $F = MS_T/MS_{ST}$ as calculated through ANOVA. Table 6 uses the example data in this analysis.

Insert Table 6 about here

This procedure of ANOVA through regression is actually using planned contrasts. When $k=2$, the omnibus and planned contrast tests are equivalent. However, when $k \geq 3$, the contrast variables defined in the first step of this procedure provide opportunities to consider specific hypotheses concerning the treatment levels, or to further partition the explained variance. These contrast variables can be designed to test mean group differences, trend analyses, or other hypotheses of interest. To test the hypothesis of contrast i, compute

$$F = \frac{r_{yi}^2/1}{(1 - R^2)/(k-1)(n-1)}$$

which is equivalent to $$F = MS_{Ti}/MS_{SI}$$

Caution needs to be taken when using the omnibus F test with repeated measures designs. To test the hypotheses of main effects or interactions using the F statistic, three assumptions must be met: 1) the k observations for each subject are drawn from a multivariate normal distribution, 2) subjects are independently sampled, and 3) the variance-covariance matrix for the k levels is spherical, or the sampling variances for all pairwise differences among means are equal. The third assumption is known as sphericity, or circularity. Both the multivariate normal and the sphericity assumptions will always be false (except if there are only two levels, when sphericity will be trivial). The F test is robust to violations of the multivariate normal assumption, but not to the sphericity assumption (Lewis, 1993). Thus, researchers must consider the extent to which sphericity is violated in their data when dealing with factors with more than two levels. In fact, Huck and

Cormier (1996) recommend that if researchers have not investigated the sphericity assumption, they should disregard all of their inferential claims.

There are several statistical tests that researchers may use to test the sphericity assumption. However, it has been shown that these tests are highly sensitive to departures from multivariate normality and from their respective null hypotheses (Barcikowski & Robey, 1984; Stevens, 1996). Box (1954) researched the effects of sphericity assumption violations on the F test. When the sphericity assumption is violated, the Type I error rate is underestimated. Box, in this situation, found that under the null hypothesis of no mean difference among the repeated measures, the sampling distribution of the standard F statistic can be approximated by an F-distribution with reduced degrees of freedom for error. The amount of reduction is dependent on the severity of the sphericity assumption violation which is estimated by ε.

Geisser and Greenhouse (1958) found the lower bound for ε which occurs when all factors have only two levels and, thus, sphericity is a trivial assumption. By using the lower bounds for degrees of freedom, 1 and n-1, the F test becomes conservative. But, since the calculations are simple, this approach is useful when researchers need a quick estimation or want to check journal articles in which no correction is used (Lewis, 1993).

Consider an example from Edwards (1985) of a one-way repeated measures design: n=5 rats were tested in k=4 trials through a maze where the number of errors each rat made on each trial was counted. For the standard F test, the degrees of

freedom are k-1=3 and (k-1)(n-1)=(3)(4)=12. With the Geisser-Greenhouse correction, the degrees of freedom are $F(1,n-1)=F(1,4)$.

A more reasonable approach when the full data set and computer software are available, would be to run the standard F test. If the result is statistically non-significant, then no further adjustment need be made since the test will only become more conservative. If the result is statistically significant, then a quick estimation based on the Geisser and Greenhouse lower bound of $F(1, n-1)$ can be made. If the result based on the most conservative test is statistically significant, then no other adjustments need be made. However, if the result is statistically non-significant, then it may be worth while to estimate ε more accurately (Huck & Cormier, 1996). Lewis (1993) and Stevens (1996) include detailed discussion along with pertinent references concerning the most appropriate estimate of ε to use. SPSS for the microcomputer will compute ε statistic if requested.

Continuing with the rats in the maze example, the observed $F=MS_T/MS_{ST}=(33.2/3)/(10.3/12)=12.89$. For the standard F test at α=0.05, the calculated F=3.49, at which the observed result is statistically significant. Using the method of checking for sphericity violations outlined above, the next step is to perform the statistical test using Geisser-Greenhouse corrected degrees of freedom. Thus, for the corrected $F(1, 4)$ at α=0.05, F=7.71. The observed F is greater than the calculated F, and, therefore, statistically significant, even when using the most conservative test. Thus, there is no need to estimate the sphericity assumption violation more accurately for this data set.

Another approach to repeated measures analyses is through using multivariate statistical techniques. This requires a paradigm shift. When considering the univariate analysis techniques, the experimental design was subjects as a random factor crossed with treatments or repeated measures as a fixed factor. To shift to the multivariate techniques, the repeated measures become a series of dependent variables and subjects are considered as replications in a single-cell design (Lewis, 1993). The most common approach is to transform the k dependent variables into k-1 linearly independent pairwise difference scores. Analysis is performed on these k-1 new dependent variables. The null hypothesis that is most often tested in this situation is that the difference scores have population means of zero, using an F transformation of Hotelling's T^2 (Lewis, 1993; Stevens, 1996).

There are advantages and disadvantages to using the multivariate approach. The multivariate approach does not require the sphericity assumption. However, researchers have not come to an agreement as to the best multivariate approach to take when considering power and robustness against assumption violations. There are serious concerns about power when the number of subjects is less than or equal to the degrees of freedom for a repeated measures main effect or interaction; in fact, the test statistic could not be computed. When the number of subjects is greater than, but still close to the degrees of freedom, the test has little power. But, power increases rapidly as the number of subjects increases (Lewis, 1993; Stevens, 1996).

In general, it is recommended that both the univariate and the multivariate approaches be run since the two approaches evaluate different aspects of the data. The only safeguard if this approach is taken is to decrease the α for each approach by

half, in order to control for experiment-wise Type I error (Barcikowski & Robey, 1984; Lewis, 1993; Stevens, 1996).

Summary

Repeated measures designs offer researchers ways to test research hypotheses by controlling for subject variance. Through these designs, greater statistical power relative to sample size is achieved. However, threats to internal validity such as carryover or practice effects need to be taken into consideration. Once data are gathered, researchers have several options for data analysis. If univariate statistical methods are used, omnibus tests can be used but must be evaluated for violation of the sphericity assumption, or planned comparisons can be used. Researchers may also use multivariate statistical methods or they may implement both univariate and multivariate approaches while controlling for experiment-wise Type I error.

References

Barcikowski, R.S., & Robey, R.R. (1984). Decisions in single group repeated measures analysis: Statistical tests and three computer packages. *The American Statistician*, 38, 148-150.

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *The Annals of Mathematical Statistics*, 25, 484-498.

Edwards, A.L. (1985). Multiple regression and the analysis of variance and covariance (2nd ed.). New York: Freeman.

Geisser, S. & Greenhouse, S.W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *The Annals of Mathematical Statistics*, 29, 885-891.

Hertzog, C., & Rovine, M. (1985). Repeated-measures analysis of variance in developmental research: Selected issues. *Child Development*, 56, 787-809.

Huck, S.W., & Cormier, W.H. (1996). Reading statistics and research (2nd ed.). New York: Harper Collins

Lewis, C. (1993). Analyzing means from repeated measures data. In G. Keren & C. Lewis (eds.), A Handbook for data analysis in the behavioral sciences (pp. 73-94). Hillsdale, NJ: Erlbaum.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Hillsdale, NJ: Erlbaum.

Table 1
Data Matrix for a General One-Way
Repeated Measures Design

	Treatments (k)				
Subjects	Y1	Y2	. . .	Yk	Σ
1	y11	y12		y1k	y1.
2	y21	y22		y2k	y2.
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
n	yn1	yn2		ynk	yn.
Σ	y.1	y.2	. . .	y.k	y..

Table 2
Data Matrix for an Example One-Way
Repeated Measures Design
for n=10 Subjects under k=2 Treatments

	Treatments (2)		
Subjects	T1	T2	Σ
1	5	3	8
2	8	4	12
3	5	6	11
4	6	5	11
5	10	6	16
6	6	4	10
7	8	8	16
8	7	5	12
9	8	6	14
10	9	3	12
Σ	72	50	122

Table 3
Summary of the Analysis of Variance
for a General One-Way Repeated Measures Design

Source	Sum of Squares	df	Mean Square	F	ES
Subjects	SSs	n-1	SSs/(n-1)		
Treatments	SSt	k-1	SSt/(k-1)	MSt/MSst	SSt/TSS
S x T	SSst	(k-1)(n-1)	SSst/(k-1)(n-1)		
Total	TSS	kn-1			

Table 4
Summary of the Analysis of Variance
for the Example One-Way Repeated Measures Design

Source	Sum of Squares	df	Mean Square	F	ES
Subjects	28.8	9	3.200		
Treatments	24.2	1	24.200	11.59	0.34
S x T	18.8	9	2.089		
Total	71.8	19			

Table 5
Mutually Orthogonal Coded Vectors for the
Example One-Way Repeated Measures Design

Subjects	X1	X2	Y
1	1	8	5
2	1	12	8
3	1	11	5
4	1	11	6
5	1	16	10
6	1	10	6
7	1	16	8
8	1	12	7
9	1	14	8
10	1	12	9
1	-1	8	3
2	-1	12	4
3	-1	11	6
4	-1	11	5
5	-1	16	6
6	-1	10	4
7	-1	16	8
8	-1	12	5
9	-1	14	6
10	-1	12	3

Table 6
Regression Analysis Using Example Data
of a One-Way Repeated Measures Design

$$r_{y1}^2 = \frac{(\Sigma X_1 Y)^2}{\Sigma X_1^2 \Sigma Y^2}$$

$$r_{y1}^2 = \frac{(22)^2}{(20)(71.8)} = .33705$$

$$r_{y2}^2 = \frac{(57.6)^2}{(115.2)(71.8)} = .40111$$

$$R^2 = \Sigma r_{y1}^2$$

$$R^2 = .33705 + .40111 = .73816$$

$$R_T^2 = \Sigma_{(i=1 \text{ to } k-1)} r_{y1}^2$$

$$R_T^2 = r_{y1}^2 = .33705$$

Note: Compare this effect size to the one found through the ANOVA summary table.

$$F = \frac{R_T^2/(k-1)}{(1 - R^2)/(k-1)(n-1)}$$

$$F = \frac{.33705/1}{.26184/9} = 11.58$$

with degrees of freedom 1 and 9

Note: This test statistic is equivalent to the F as calculated through ANOVA.